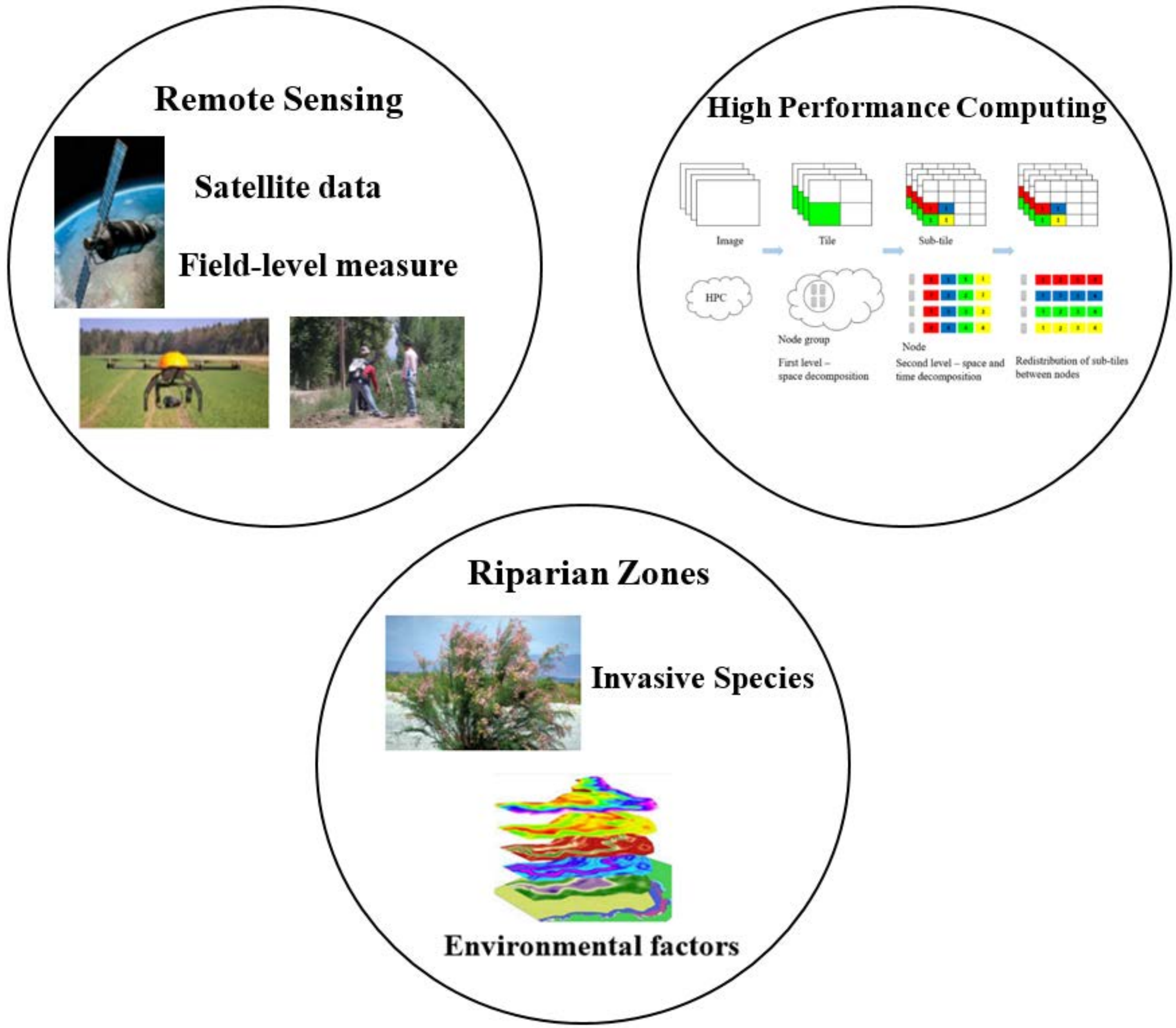
Continental-scale remote monitoring of invasive species dynamics through petascale high performance computing system

Chunyuan Diao

Assistant Professor
Department of Geography and GIScience
University of Illinois at Urbana-Champaign

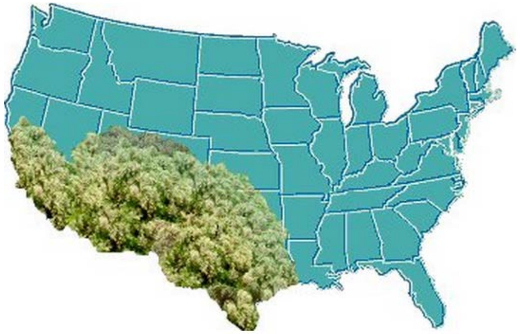
I use Blue Waters to prototype a parallel computational framework to handle massive amount of satellite data for large-scale invasive species monitoring



Introduction

- **Saltcedar is an exotic shrub species invading riparian zones of the United States**

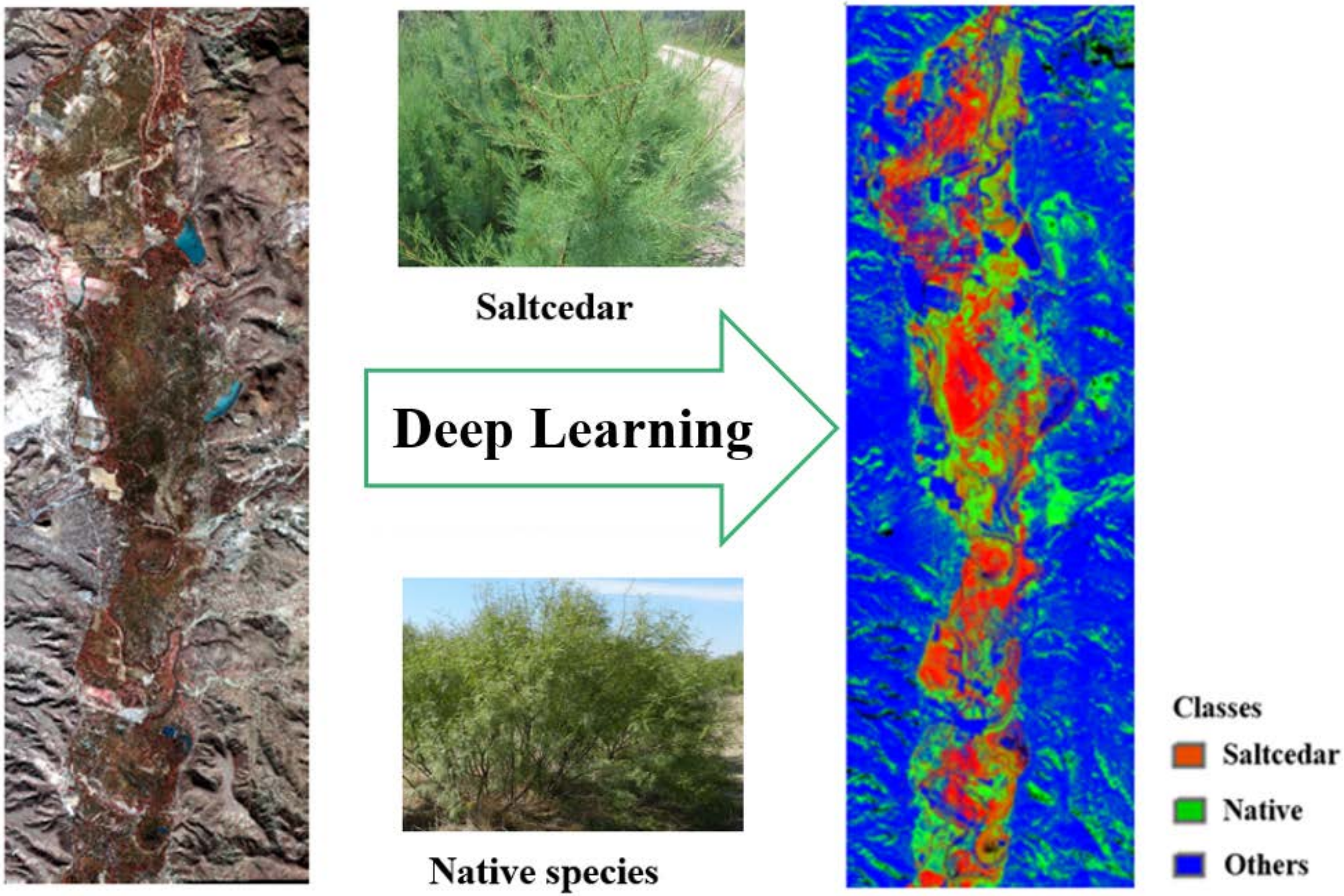
  - Alter stream hydrology
  - Increase soil salinity
  - Degrade habitats for native species



Annual economic losses from saltcedar in the US are estimated to be $133-285 million

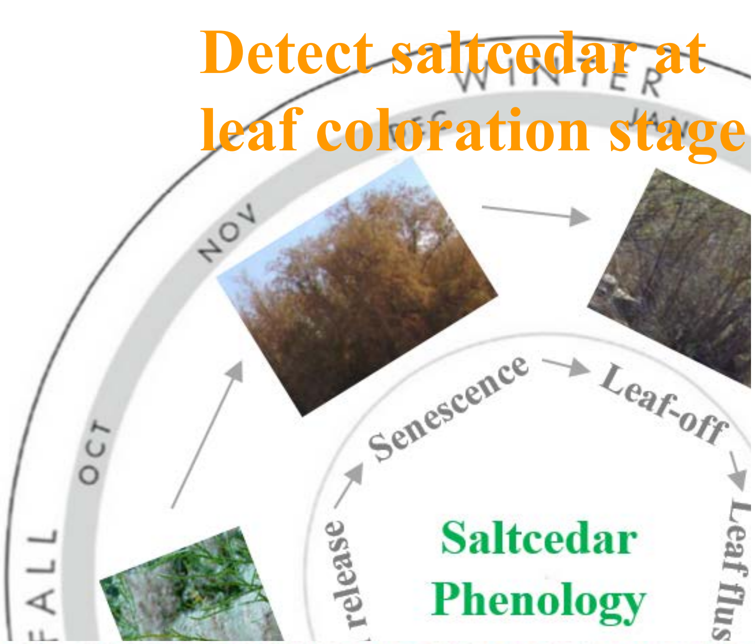
Remote Sensing

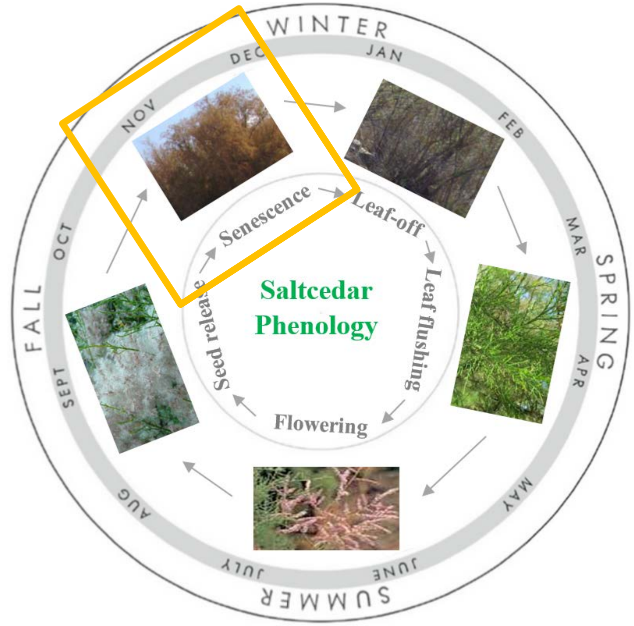
- **Machine/deep learning to map saltcedar distribution**

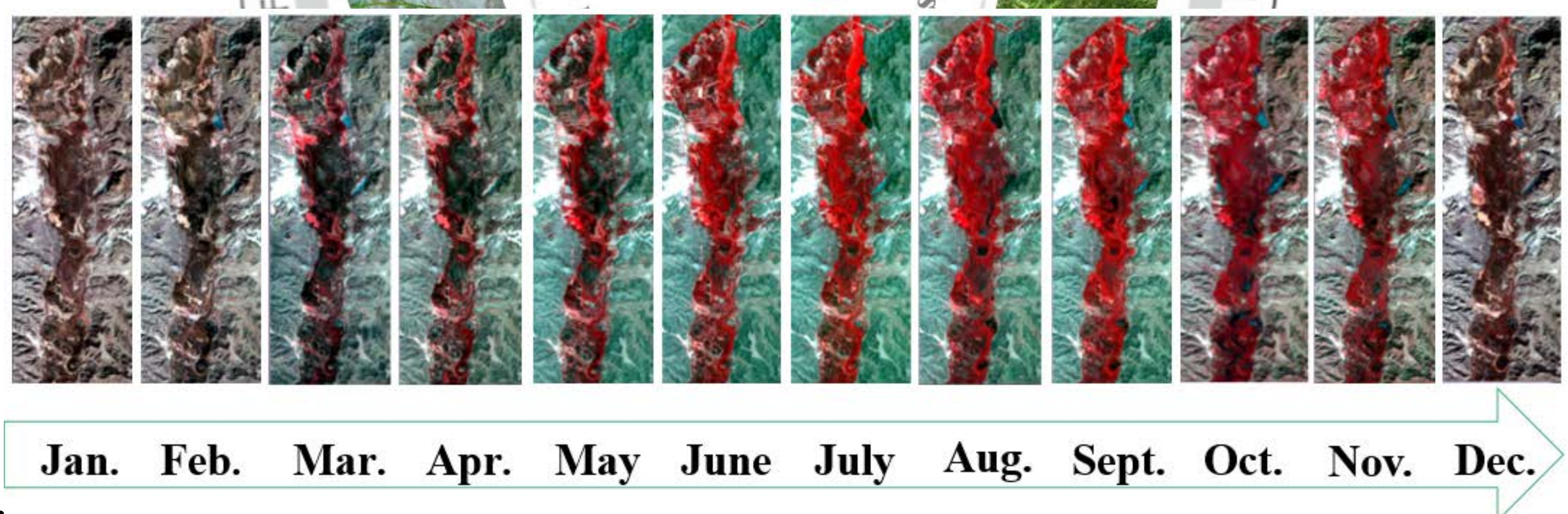


Saltcedar Phenology

Detect saltcedar at leaf coloration stage







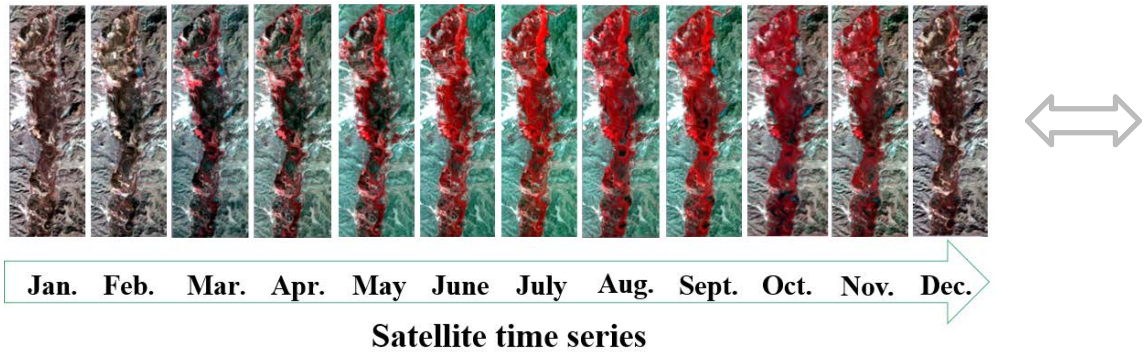
Satellite time series

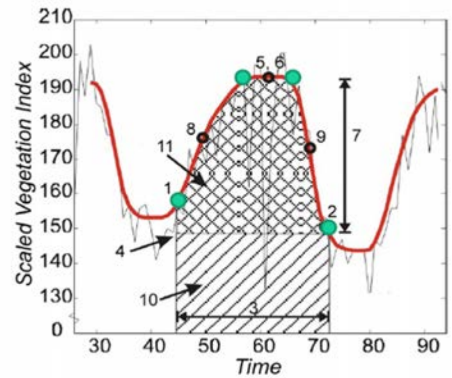
Diao,

detec 0-71.

Challenges

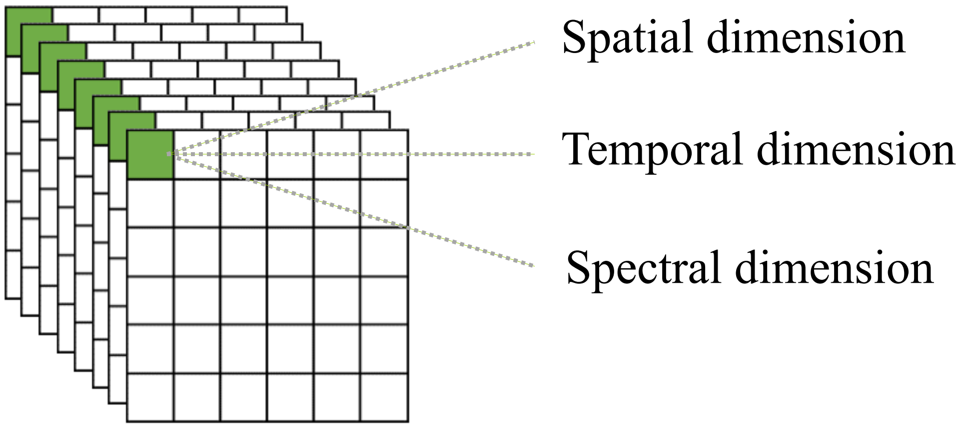
1) **Leaf coloration timing cannot be predicted using current phenological models**





Conventional phenological models

2) **Massive volume of satellite data cannot be adequately handled by traditional remote sensing systems**



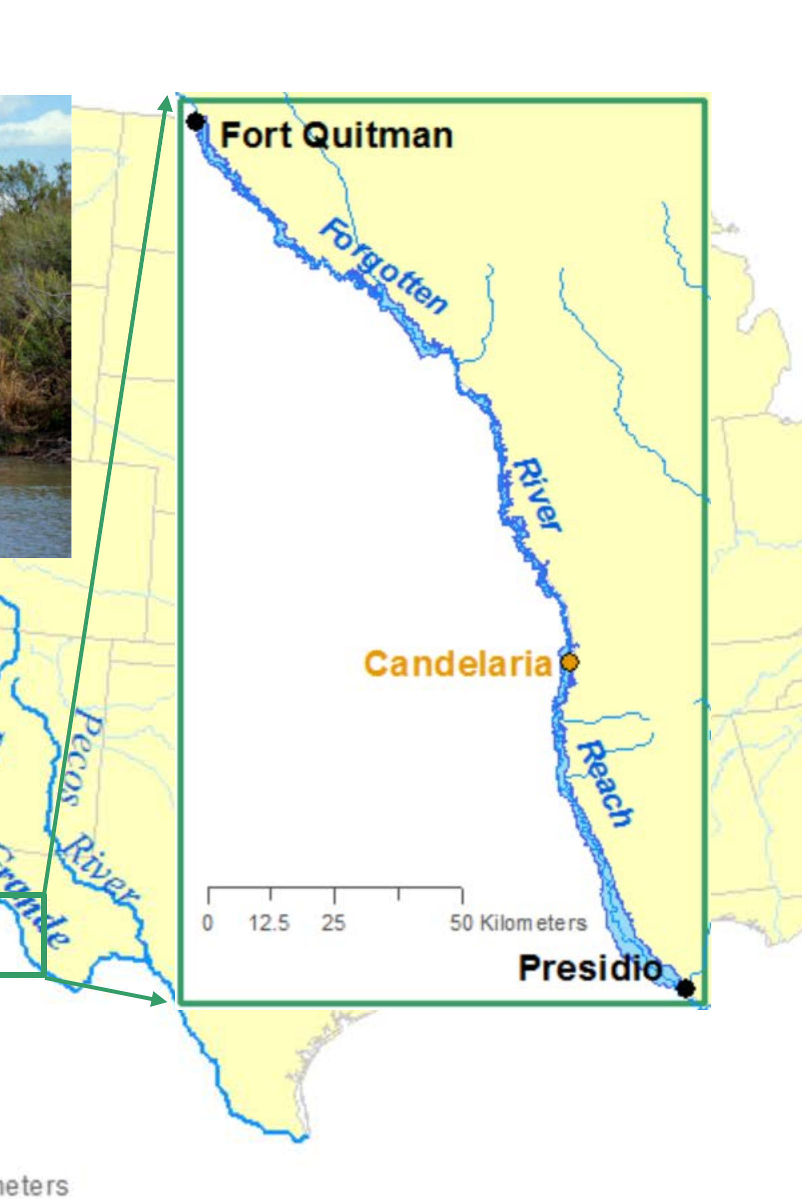
Objective

- **Develop a parallel computational framework to model the spatio-temporal dynamics of saltcedar over the past 40 years**

1) Develop computational algorithms that can model the leaf coloration stage of invasive saltcedar using satellite time series

2) Devise a high-performance parallel system to prototype the data- and compute-intensive satellite invasive species monitoring system

Study Site



Candelaria, TX

Parallel computational framework

1. **Leaf coloration computational algorithms**

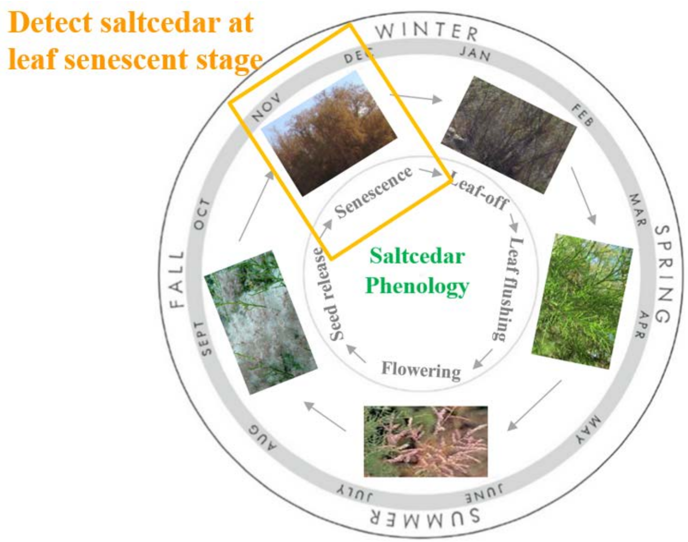
2. High-performance parallel system

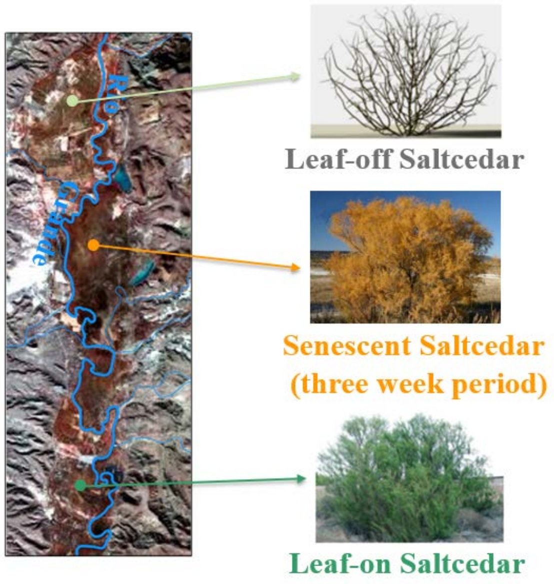
Computational algorithms

- **To model and predict the timing of saltcedar coloration**

1. Multiyear Spectral Angle Clustering Model – sparse satellite time series (Diao and Wang, *Remote Sensing of Environment*, 2018)

2. Pheno-network Model – dense satellite time series (Diao, *Remote Sensing of Environment*, 2019)



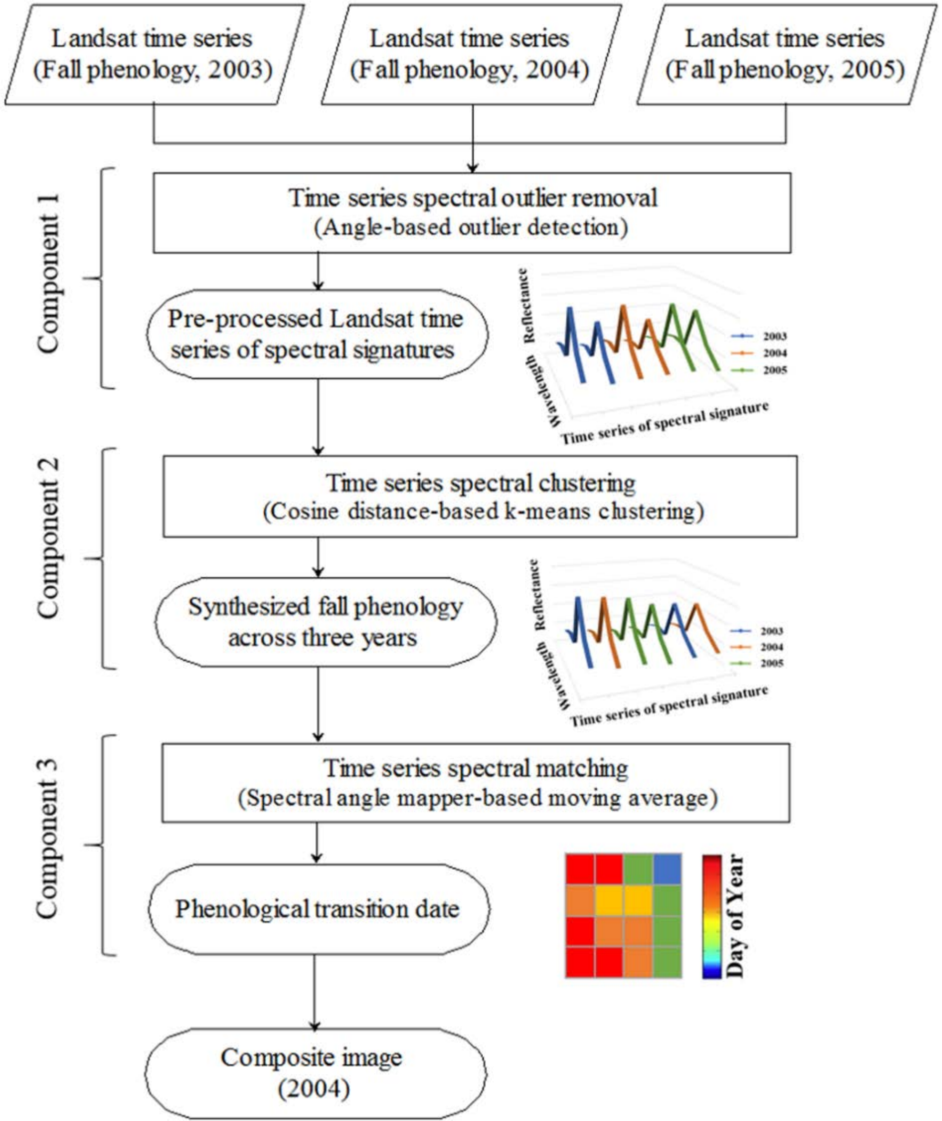


Diao, C. and L. Wang. (2018). Landsat time series-based multiyear spectral angle clustering (MSAC) model to monitor the inter-annual leaf senescence of exotic saltcedar. *Remote Sensing of Environment*, 209, 581-593.

Diao, C. (2019). Complex network-based time series remote sensing model in monitoring the fall foliage transition date for peak coloration. *Remote Sensing of Environment*, 229, 179-192.

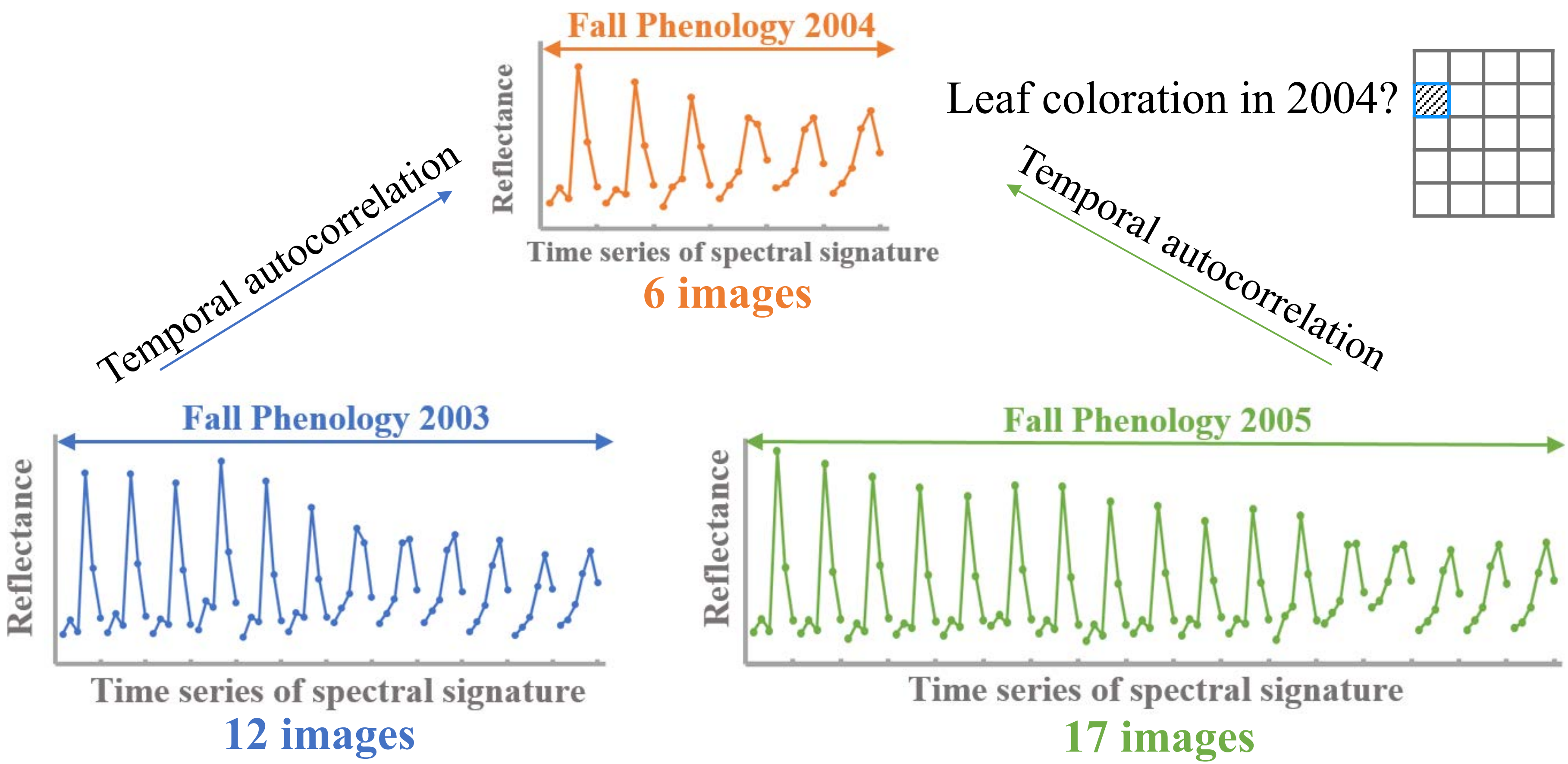
Multiyear Spectral Angle Clustering Model

- **To model the timing of saltcedar coloration with sparse time series**



Multiyear Spectral Angle Clustering Model

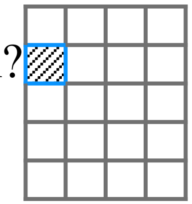
- **To model the timing of saltcedar coloration with sparse time series**



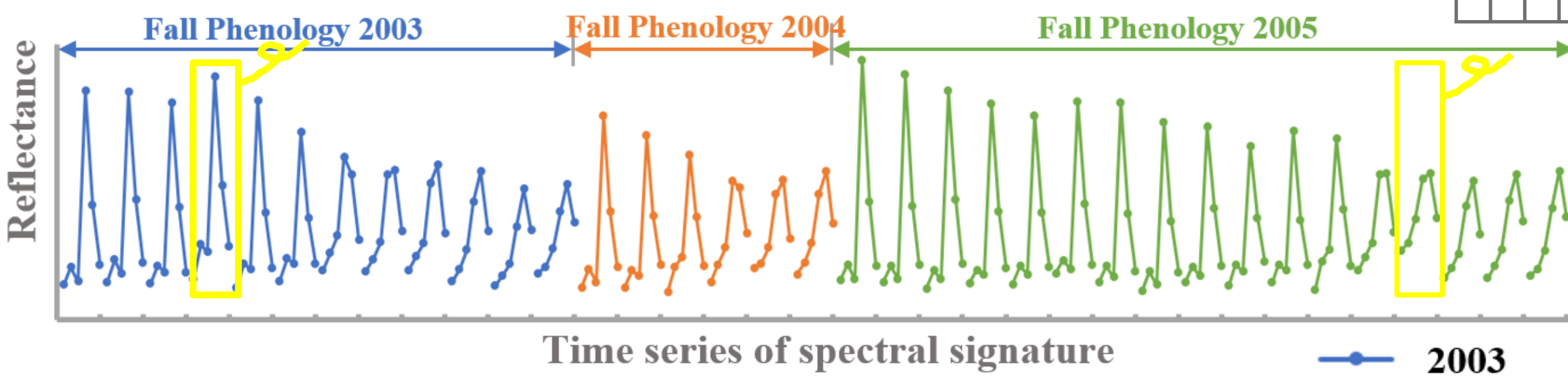
Multiyear Spectral Angle Clustering Model

1) Time series spectral outlier removal

Leaf Coloration?

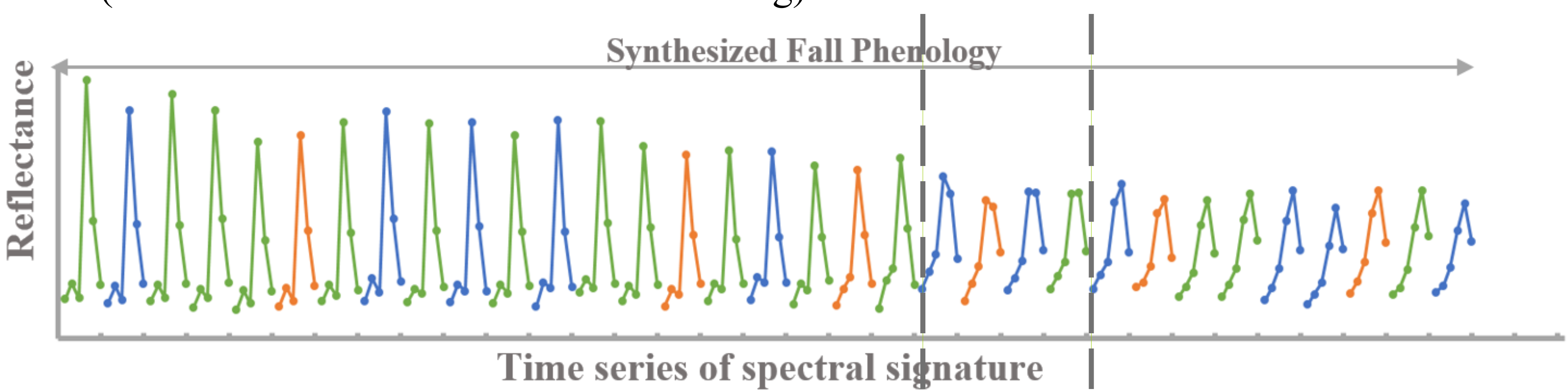


(Angle-based outlier detection method)



2) Time series spectral clustering

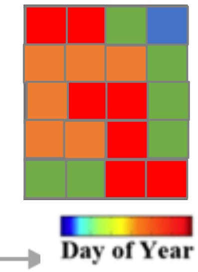
(Cosine distance-based k-means clustering)

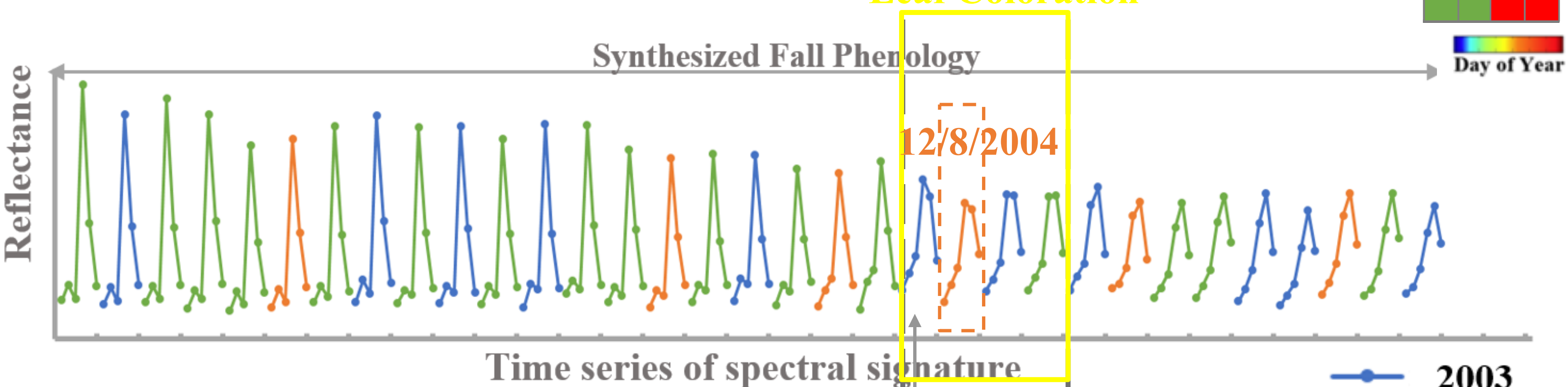


Multiyear Spectral Angle Clustering Model

3) Time series spectral matching

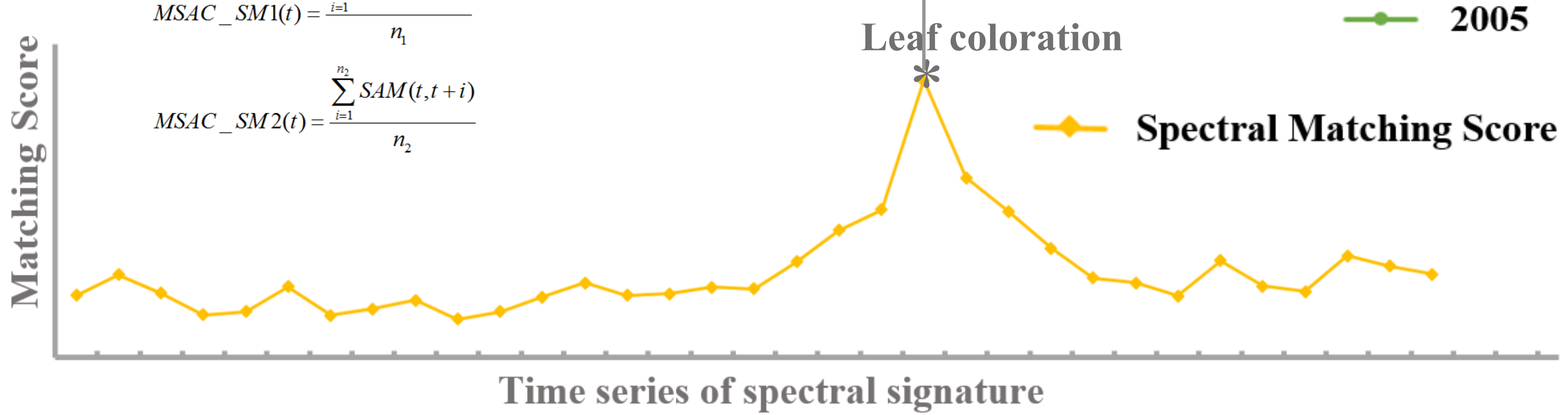
(Spectral angle mapper-based moving average method)





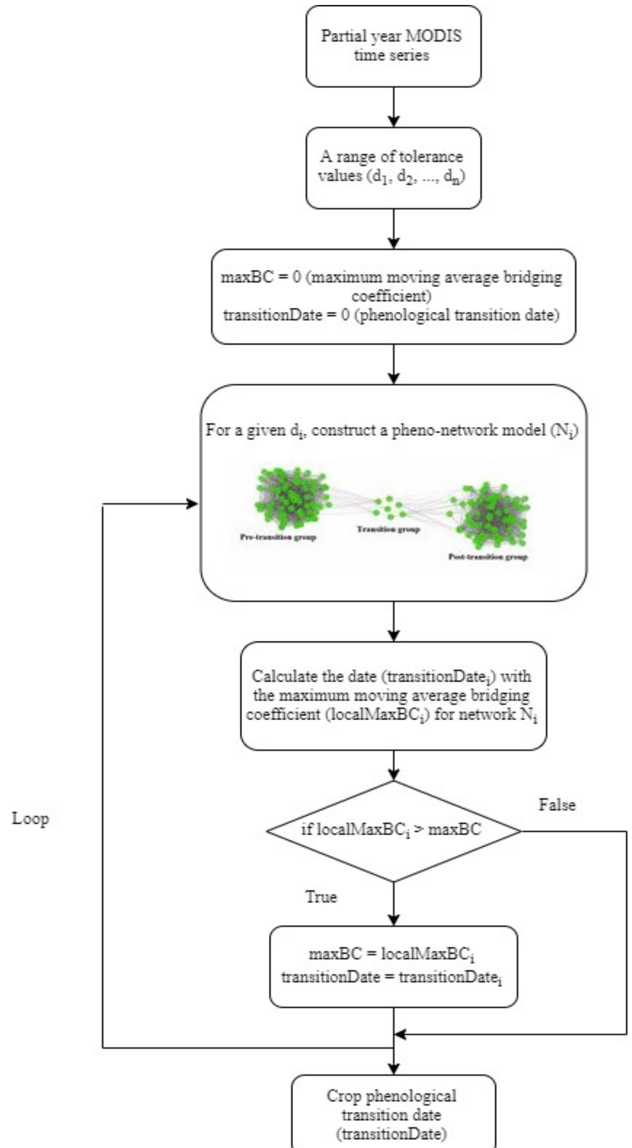
$$MSAC\_SM1(t) = \frac{\sum_{i=1}^{n_1} SAM(t, t-i)}{n_1}$$

$$MSAC\_SM2(t) = \frac{\sum_{i=1}^{n_2} SAM(t, t+i)}{n_2}$$



Pheno-network

- **To model the timing of saltcedar coloration with dense time series**

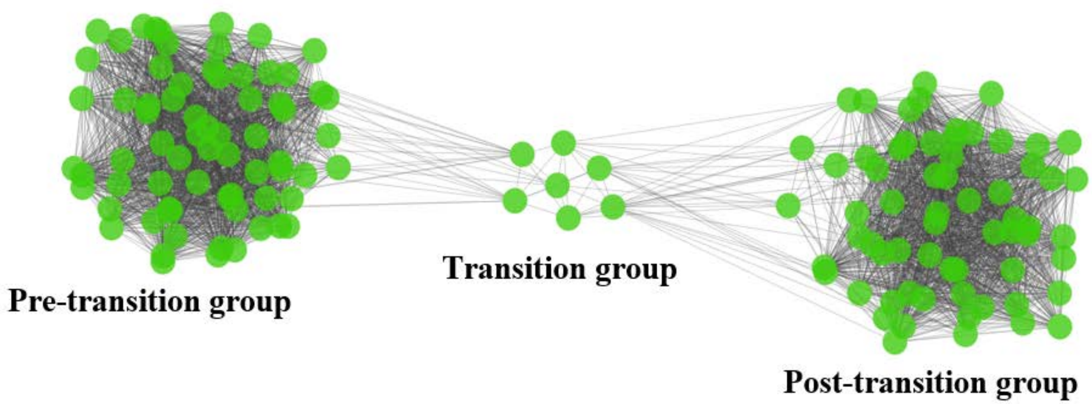


Pheno-network Model

- **Network representation of saltcedar phenological progress**

  - Node: spectral reflectance obtained on each date of the time series

  - Edge: spectral similarity between the spectral nodes



Pheno-network with three groups, namely the pre-transition, transition, and post-transition groups.

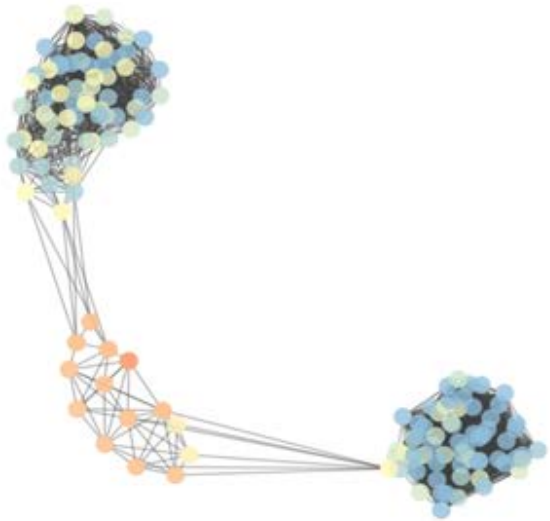
Pheno-network Model

- **Network measures of saltcedar leaf coloration**

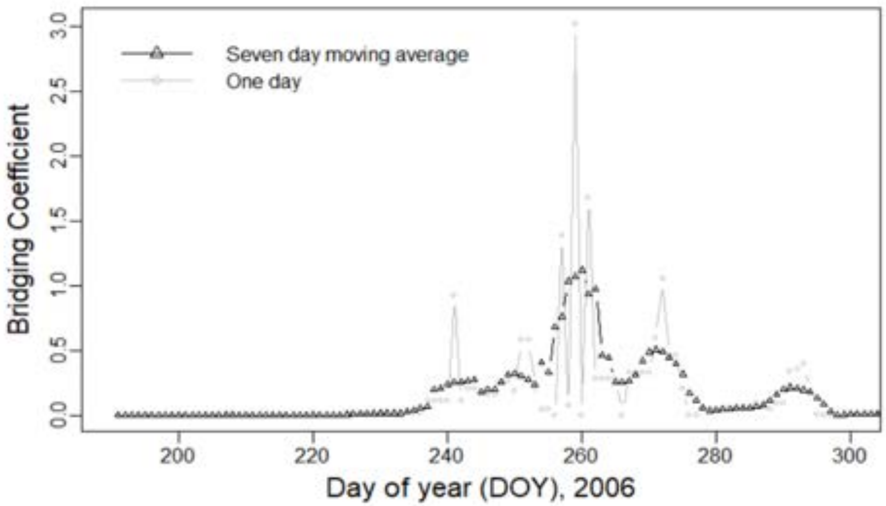
  - Betweenness Centrality: the transition node serves as the hub connecting the nodes across phenological stages

  - Clustering Coefficient: the neighbors of the transition node are sparsely connected to each other

$$\text{Bridging Coefficient} = \frac{\text{Betweenness Centrality}}{\text{Clustering Coefficient}}$$



(a) Pheno-network of saltcedar pixel in 2006



(b) Network measures of spectral nodes in the pheno-network

Composite Landsat Image

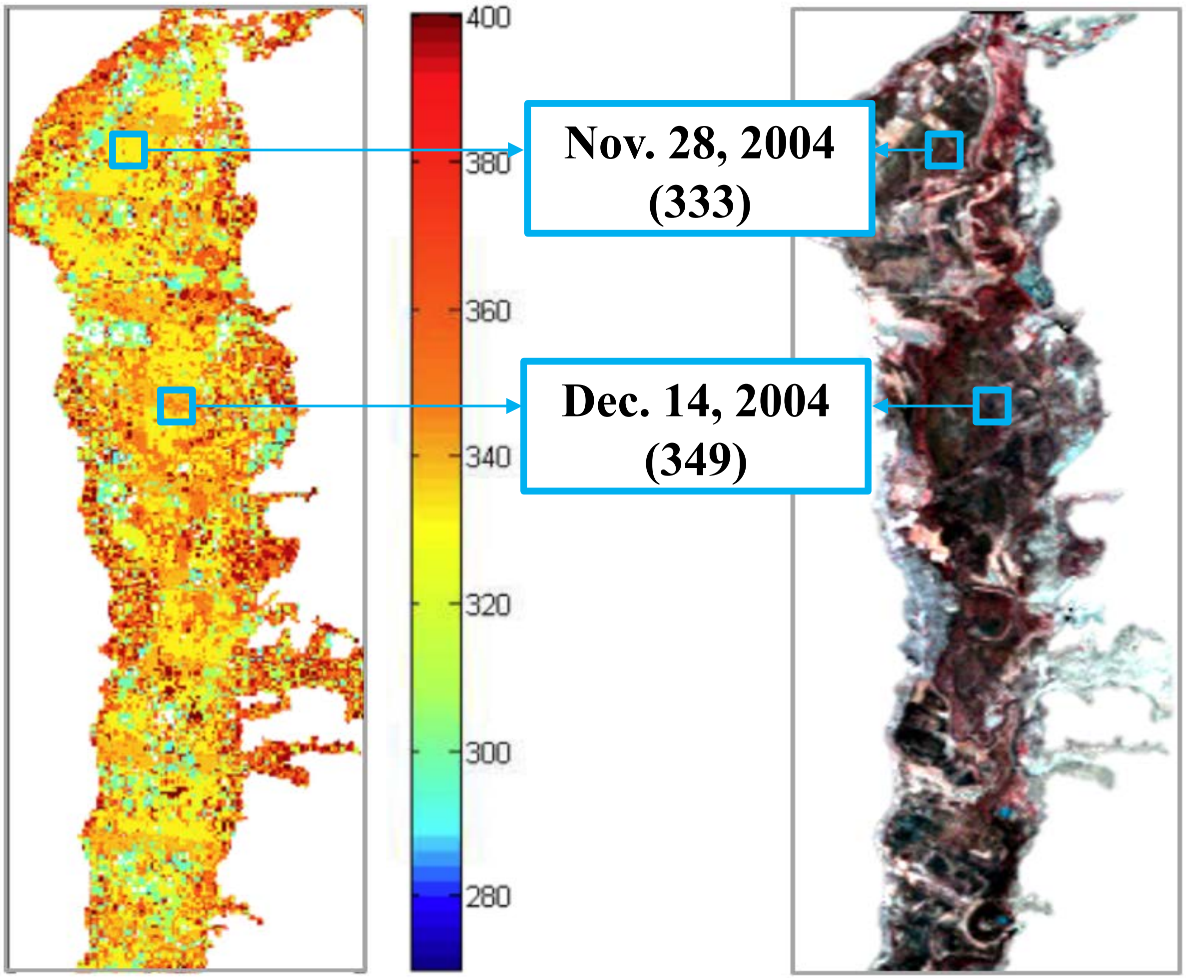


Image acquisition date at leaf senescence (2004)

Composite image (2004)

Composite Landsat Image

Overall Accuracy: 81.25%

Kappa: 0.65

Producer's Accuracy: 76%

User's Accuracy: 83%

⇕

Single Landsat Image (12/8/2004)

Overall Accuracy: 74.25%

Kappa: 0.49

Producer's Accuracy: 66%

User's Accuracy: 79%

Parallel computational framework

1. Leaf coloration computational algorithms

2. **High-performance parallel system**

Conventional remote sensing system

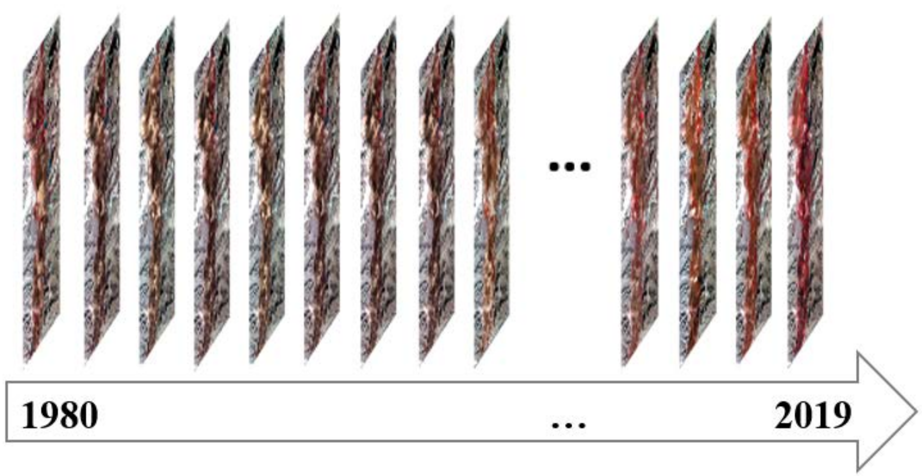
- **Conventional remote sensing systems analyze entire remote sensing imagery as a whole**

  - High memory requirements and low scalability

- **Large-scale remote sensing monitoring is challenging**

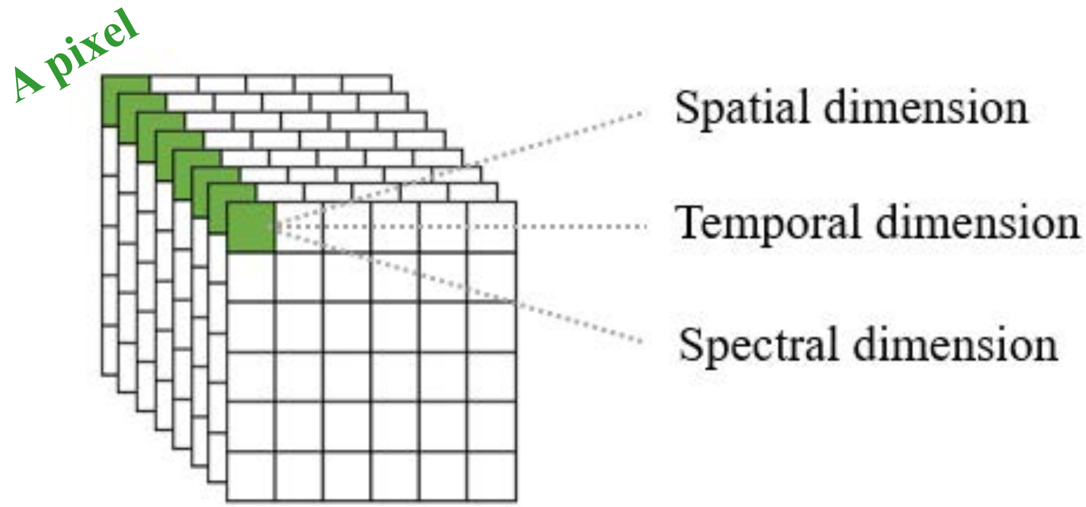
  - Massive amount of satellite imagery

  - High demands for computational resources



High performance parallel system

- **The leaf coloration algorithms are designed at the pixel level**



- **The parallel system decomposes the remote sensing imagery into a multitude of sub-tiles**

  - Reduce memory requirement

  - Optimize I/O and computation time

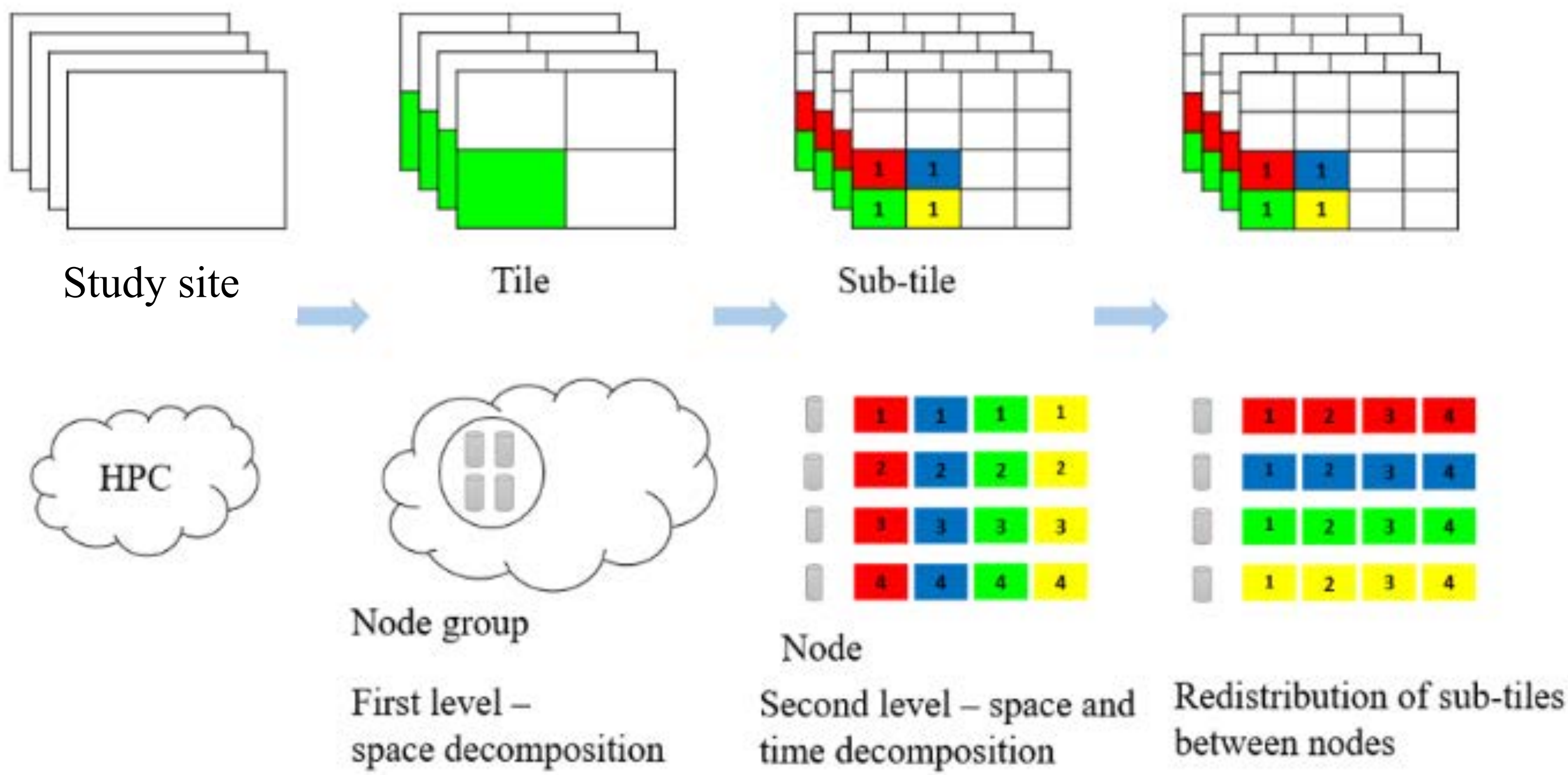
High performance parallel system

The parallel system adopts hybrid computation models

- **Node-level data distribution model – MPI**
- **Core-level computation model - OpenMP**

High performance parallel system

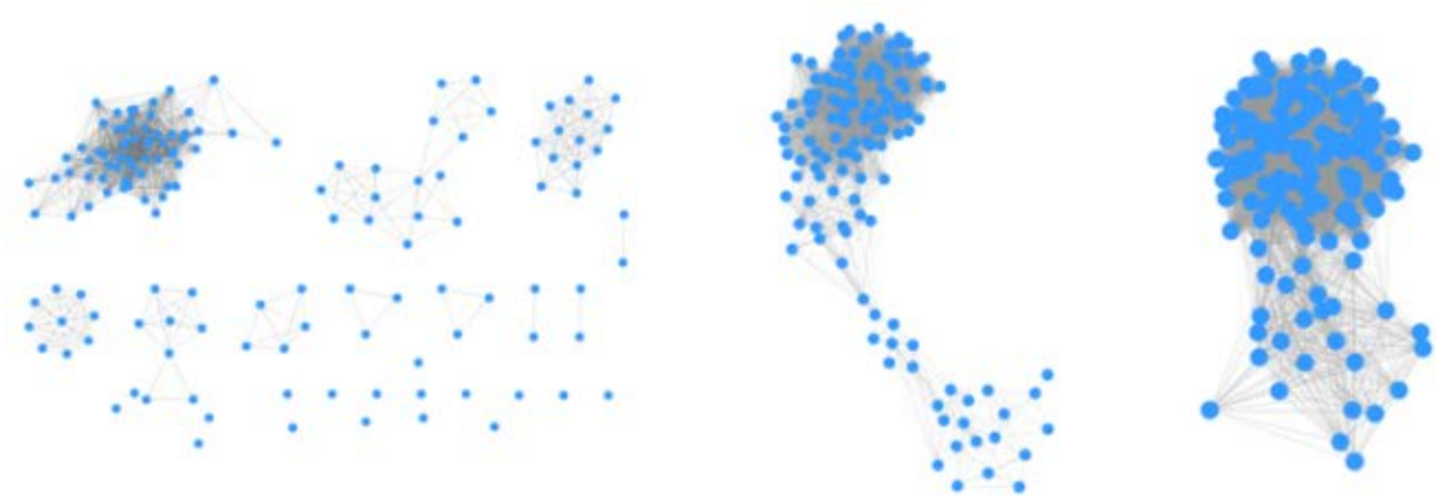
- **Node-level data distribution model**



The two-level data distribution model: massive data and I/O operations are evenly distributed among all computing nodes

High performance parallel system

- **Core-level computation model**



(a) 10th percentile of the cosine distance values

(b) 40th percentile of the cosine distance values

(c) 70th percentile of the cosine distance values

Parameter calibration in pheno-network models

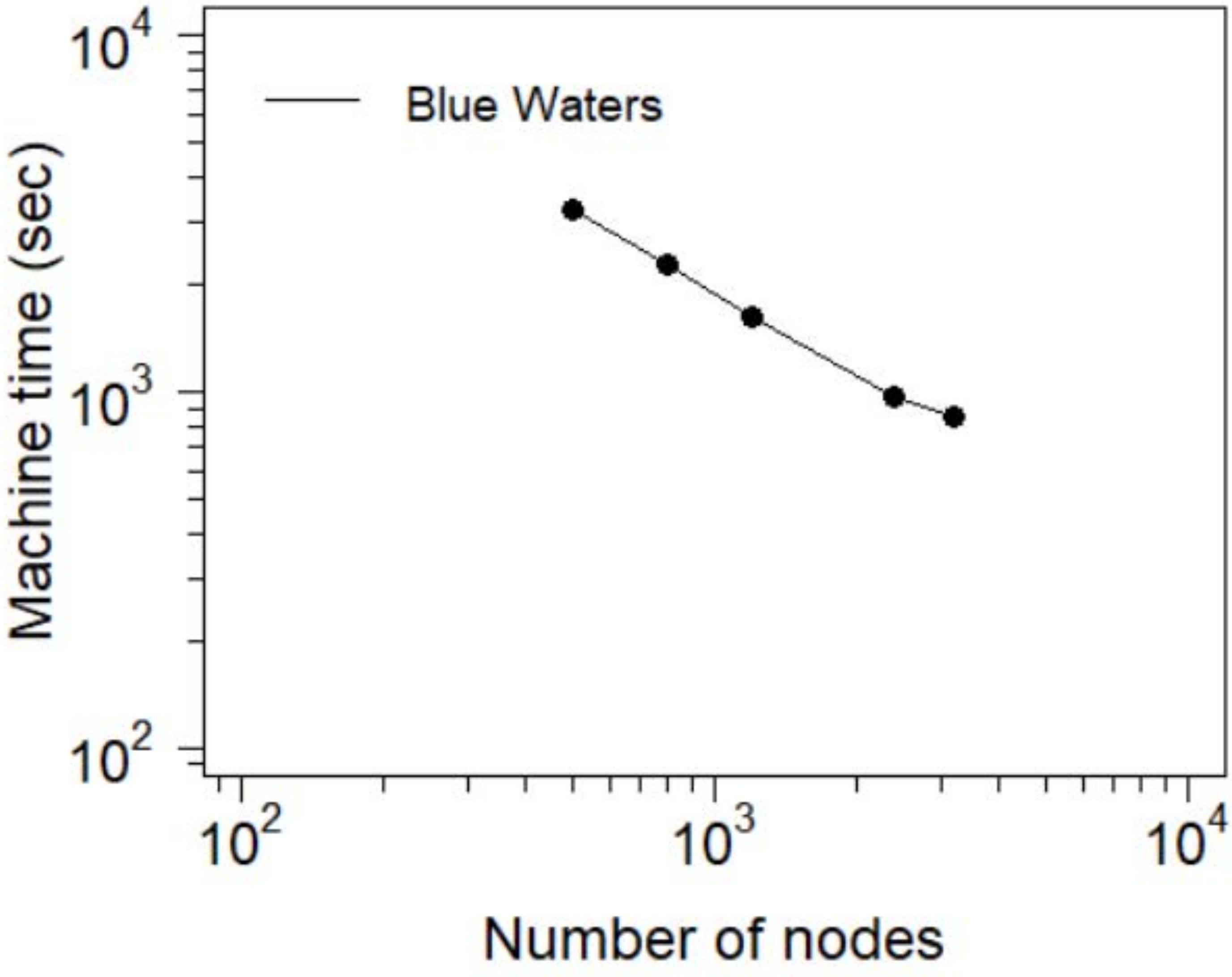
Core-level computation model increases computation efficiency while decreasing memory requirement.

Why Blue Waters?

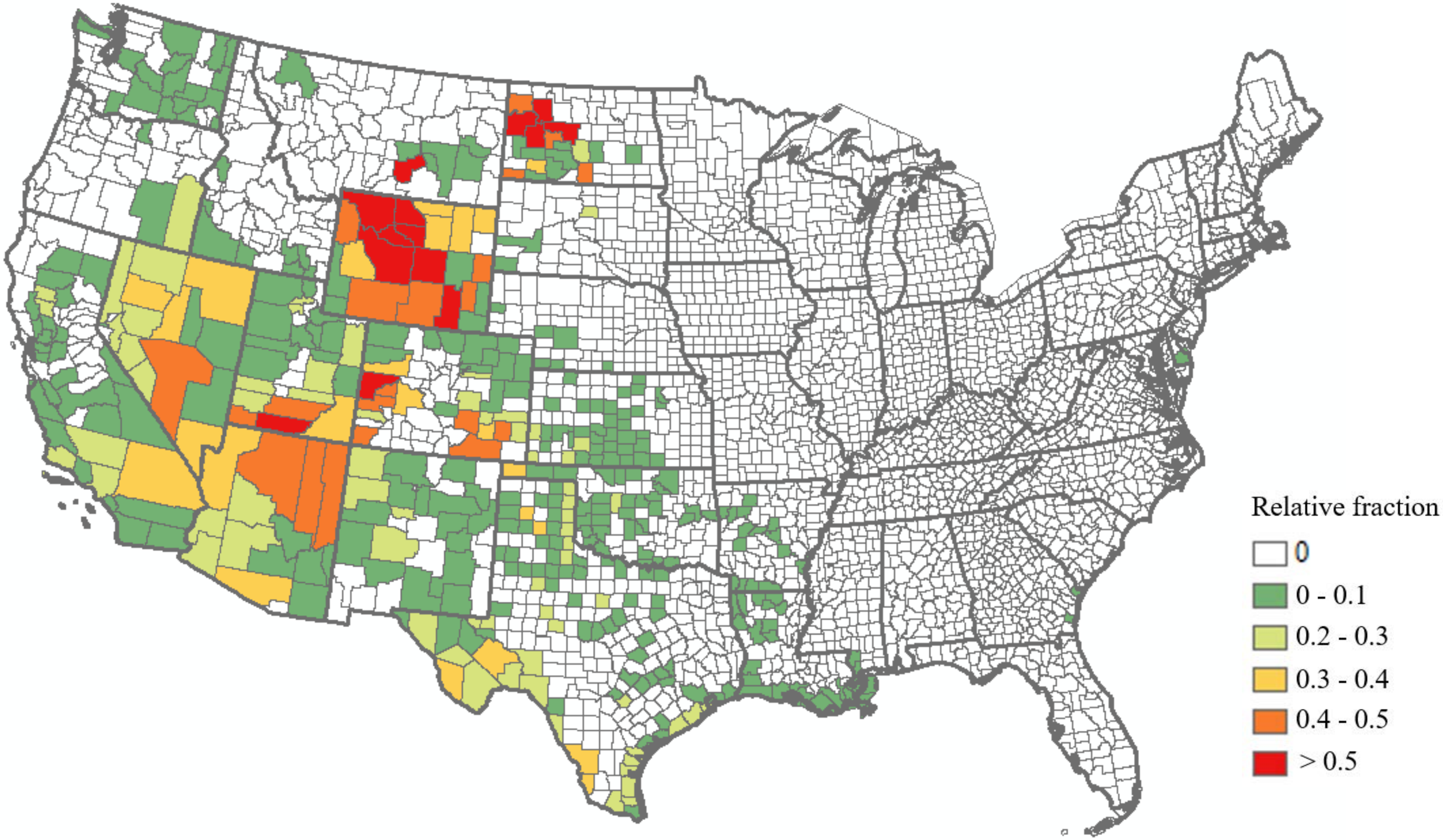
Blue Waters facilitates the processing of massive amount of satellite data with high spatial, temporal and spectral dimensions

- Large storage space
- Access to a large number of nodes
- High-speed simultaneous access to a large number of images
- Large network bandwidth to increase data distribution speed

Scalability of parallel system

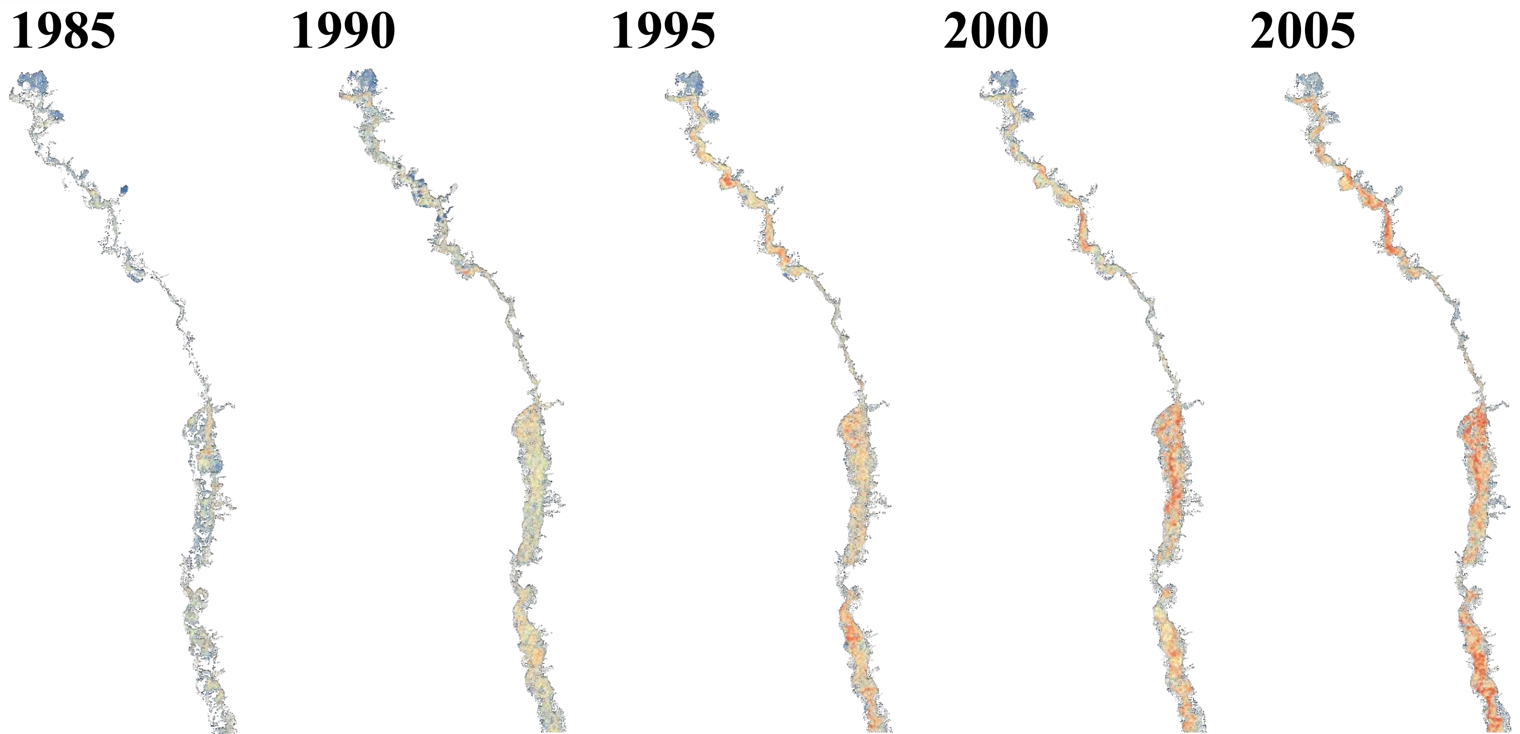


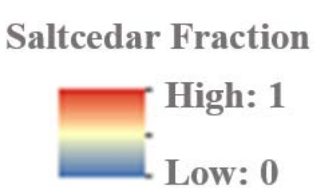
Saltcedar Distribution Map



Saltcedar distribution map in 2005

Spatio-temporal Dynamics of Saltcedar





Change in area (ha) over time for the three classes

Class	1985	1995	2005	Overall Change
Saltcedar (ha)	4497	5027	5607	+1110
Native woody riparian vegetation (ha)	2352	2201	2021	-331
Other (ha)	13743	13364	12964	-779

Conclusions

1) **The multiyear spectral angle clustering and pheno-network models can model the leaf coloration stage of invasive saltcedar**

2) **The high performance parallel system can efficiently process massive satellite time series with high scalability**

3) **Invasive saltcedar is displacing native riparian vegetation over time and space**

Reference

1. **Diao, C. (2019). Complex network-based time series remote sensing model in monitoring the fall foliage transition date for peak coloration. *Remote Sensing of Environment*, 229, 179-192.**

2. **Diao, C. and L. Wang. (2018). Landsat time series-based multiyear spectral angle clustering (MSAC) model to monitor the inter-annual leaf senescence of exotic saltcedar. *Remote Sensing of Environment*, 209, 581-593.**

3. **Diao, C. and L. Wang. (2016). Incorporating plant phenological trajectory in exotic saltcedar detection with monthly time series of Landsat imagery. *Remote Sensing of Environment*, 182, 60-71.**

Acknowledgement

- This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.
- This research is partially supported by NSF Office of Advanced Cyberinfrastructure award and Campus Research Board at the University of Illinois at Urbana-Champaign.

Thank You!